Comorphism-based Grothendieck logics

Till Mossakowski

BISS, Dept. of Computer Science, University of Bremen

Abstract. In order to obtain a semantic foundation for heterogeneous specification, we extend Diaconescu's morphism-based Grothendieck institutions to the case of comorphisms. This is not just a dualization, because we obtain more general results, especially concerning amalgamation properties. We also introduce a proof calculus for structured heterogeneous specifications and study its soundness and completeness (where amalgamation properties play a rôle for obtaining the latter).

1 Introduction and motivation

For the specification of large software systems, heterogeneous multi-logic specifications are needed, since complex problems have different aspects that are best specified in different logics. A combination of all the used logics would become too complex in many cases. Moreover, specialized languages and tools often have their strengths in particular aspects. Using heterogeneous specification, these strengths can be combined with comparably small effort.

In the literature, several approaches to heterogeneous specification have been developed [7, 8, 17, 23]. The most prominent approach is CafeOBJ with its cube of eight logics and twelve projections (formalized as institution morphisms) among them [9], having a semantics based on the notion of Grothendieck institution [8]. However, not only projections between logics, but also logic encodings (formalized as so called comorphisms) are relevant to heterogeneous specification [23, 17]. Moreover, besides these model theoretic approaches, also the need of integrating different proof calculi via "bridges" has been stressed [7]. The goal of this paper is to extend the CafeOBJ resp. Grothendieck institution approach to cover these aspects.

2 Institutions, logics, morphisms and comorphisms

We now recall several notions mentioned in the introduction. Institutions [14] capture the model theory of a logic, entailment systems [16] capture proof theory, while logics [16] combine both. Institution morphisms [14] capture the intuition that one logic is built upon, or projected onto another one, while institution comorphisms [13], also called institution representations [22] or maps of institutions [16] capture the intuition that one logic is encoded into another one. Both notions also can be extended to full logics. We begin with the components of institutions,
which are called rooms, and typically provide a satisfaction system local to some
signature.

An institution room \((S, M, \models)\) consists of

- a set of \(S\) of sentences,
- a category \(M\) of models, and
- a satisfaction relation \(\models \subseteq |M| \times S\).

Rooms are connected via corridors (which model change of notation within
one logic, as well as translations between logics).

An institution corridor \((\alpha, \beta): (S_1, M_1, \models_1) \to (S_2, M_2, \models_2)\) consists of

- a sentence translation function \(\alpha: S_1 \to S_2\), and
- a model reduction functor \(\beta: M_2 \to M_1\), such that

\[
m_2 \models_2 \alpha(\varphi_1) \iff \beta(m_2) \models_1 \varphi_1
\]

holds for each \(m_2 \in M_2\) and each \(\varphi_1 \in S_1\) (satiation condition).

Semantic entailment in an institution room is defined as usual: for \(\Gamma \subseteq S, \varphi \in S\), we write \(\Gamma \models \varphi\), if all models satisfying \(\Gamma\) also satisfy \(\varphi\).

A logic room \((S, M, \models, \vdash)\) is an institution room \((S, M, \models)\) equipped with an
entailment relation \(\vdash \subseteq \mathcal{P}(S) \times S\), such that the following conditions are satisfied:

1. reflexivity: for any \(\varphi \in S\), \(\{\varphi\} \vdash \varphi\),
2. monotonicity: if \(\Gamma \vdash \varphi\) and \(\Gamma' \supseteq \Gamma\) then \(\Gamma' \vdash \varphi\),
3. transitivity: if \(\Gamma \vdash \varphi_i\), for \(i \in I\), and \(\Gamma \cup \{\varphi_i \mid i \in I\} \vdash \psi\), then \(\Gamma \vdash \psi\),
4. soundness: for any \(\Gamma \subseteq S\) and \(\varphi \in S\),

\[
\Gamma \vdash \varphi \text{ implies } \Gamma \models \varphi.
\]

A logic room will be called complete if, in addition, the converse of the above
implication holds.

A logic corridor \((\alpha, \beta): (S_1, M_1, \models_1, \vdash_1) \to (S_2, M_2, \models_2, \vdash_2)\) is an institution corridor
\((\alpha, \beta): (S_1, M_1, \models_1) \to (S_2, M_2, \models_2)\) such that if \(\Gamma \vdash_1 \varphi\), then \(\alpha(\Gamma) \vdash_2 \alpha(\varphi)\) \((\vdash \text{-translation})\). Together with obvious notions of composition and identity, this
gives us categories \textbf{InsRoom} and \textbf{LogRoom}.

Generally, sentences and models depend on a given vocabulary of non-logical
symbols provided by a signature. Therefore, an institution is a functor \(I: \text{Sign} \to \text{InsRoom}\) (where \text{Sign} is called the category of signatures), and a logic is a
functor \(I: \text{Sign} \to \text{LogRoom}\). We will also use the more standard notation
\((\text{Sen}'(\Sigma), \text{Mod}'(\Sigma), \models_{\Sigma})\) for the institution room \(I(\Sigma)\). \text{Mod}' and \text{Sen}' can
easily be seen to be functors, and we arrive at the standard definition of
institutions as a quadruple \((\text{Sign}, \text{Sen}, \text{Mod}, \models)\). We will freely use this notation
whenever needed, and also write \(m|_{\sigma}\) for \(\text{Mod}(\sigma)(m)\).

For the morphisms between institutions and logics, there are two obvious choices:
Given institutions \(I_1: \text{Sign} \to \text{InsRoom}\) and \(I_2: \text{Sign} \to \text{InsRoom}\),
an institution morphism \((\Phi, \mu): I_1 \to I_2\) consists of a functor \(\Phi: \text{Sign} \to \text{Sign}\) and
a natural transformation \(\mu: I_2 \circ \Phi \Rightarrow I_1\). In contrast, an institution
comorphism \((\Phi, \rho): I_1 \to I_2\) consists of a functor \(\Phi: \text{Sign} \to \text{Sign}\) and a natural

transformation \( \rho : I_1 \rightarrow I_2 \circ \Psi \). Together with obvious compositions and identities, this gives us categories \textbf{Ins} and \textbf{colIns}. Logic morphisms and comorphisms are defined analogously, leading to categories \textbf{Log} and \textbf{coLog}.

Given institution morphisms \((\Psi, \mu): I_1 \rightarrow I_2 \) and \((\Psi', \mu') : I_1 \rightarrow I_2 \), an \textit{institution morphism transformation} \( \theta : (\Psi, \mu) \rightarrow (\Psi', \mu') \) is just a natural transformation \( \theta : \Psi \rightarrow \Psi' \) such that \( \mu = \mu' \circ (I_2 \cdot \theta) \). \(^1\) Similarly, given institution comorphisms \((\Phi, \rho) : I_1 \rightarrow I_2 \) and \((\Phi', \rho') : I_1 \rightarrow I_2 \), an \textit{institution comorphism transformation} \( \theta : (\Phi, \rho) \rightarrow (\Phi', \rho') \) is a natural transformation \( \theta : \Phi \rightarrow \Phi' \) such that \( (I_2 \cdot \theta) \circ \rho = \rho' \).

The corresponding notions for logics are entirely analogous. This shows that \textbf{Ins}, \textbf{colIns}, \textbf{Log} and \textbf{coLog} are indeed 2-categories. In the sequel, most of the definitions and results hold for both institutions and logics, although we will not repeat this every time (indeed, most constructions work for an arbitrary category of rooms).

![Diagram](https://example.com/diagram.png)

**Fig. 1.** Some institutions, morphisms, comorphisms and comorphism transformations

Consider, for example, the institutions in Fig. 1. \textit{PROP} is propositional logic, \textit{FOL} is first-order logic, \textit{HOL} is higher-order logic, both with equality. 
\textit{CASL} \cite{19} extends \textit{FOL} with partiality, subsorting and generation constraints (some form of induction). \textit{CASL-LTL} \cite{20} is an extension of \textit{CASL} with a CTL-like labeled transition logic. We also have included \textit{CTL} itself, a temporal logic featuring both temporal modalities and path quantifiers, with model-checkers available. \textit{SB-CASL} \cite{4} is an extension of \textit{CASL} that follows the abstract state machine paradigm, where states correspond to algebras. There are obvious projection morphisms (denoted by dotted arrows), that always come in pair with inclusion comorphisms (denoted as solid arrows). Generally, in heterogeneous specification, these morphisms can be simulated by the corresponding comor-

\(^1\) In \cite{8}, a modification between \( \mu \) and \( \mu' \circ (I_2 \cdot \theta) \) is allowed, using a 2-category of rooms, but we do not think that this extra generality is of much practical use.
phisms\(^2\). Some of these logics are represented via more complex comorphisms\(^3\) in \(HOL^-\), see [17]. Finally, comorphism transformations are denoted by double arrows. They are just inclusions here. E.g. we can go from \(FOL^-\) to \(HOL^-\) either directly or via CASL. Now the lower right double arrow tells us that the difference between these two ways can be captured by a family of signature morphisms in \(HOL^-\). Later on, this will enable us to identify some signature morphisms in the Grothendieck construction, leading to a better applicability of proof rules in the calculus for heterogeneous development graphs introduced below.

3 Grothendieck institutions and logics

Given an index 2-category \(Ind\), a 2-indexed institution is a 2-functor \(I: Ind^* \to \text{Ins}_2\) into the 2-category of institutions, institution morphisms and institution morphism transformations. In cases where the 2-categorical structure is not needed, we omit the prefix “2.”

Similarly, a 2-indexed coinstitution is a 2-functor \(I: Ind^* \to \text{coIns}\).

The Grothendieck construction for indexed institution has been described in [8]; we develop its dual here. In an indexed coinstitution \(I\), we use the notation \(I^i\) for \(I(i)\), \((\Phi^i, p^i)\) for the comorphism \(I(d)\) as so on.

**Definition 3.1.** Given a 2-indexed coinstitution \(I: Ind^* \to \text{coIns}\), define the Grothendieck institution \(I^\#\) as follows:

- signatures in \(I^\#\) are pairs \((\Sigma, i)\), where \(i \in [Ind]\) and \(\Sigma\) a signature in \(I^i\),
- signature morphisms \((\sigma, d): (\Sigma_1, i) \to (\Sigma_2, j)\) consist of a morphism \(d: j \to i\) in \(Ind\) and a signature morphism \(\sigma: \Phi^d(\Sigma_1) \to \Sigma_2\),
- composition is given by \((\sigma_2, d_2) \circ (\sigma_1, d_1) = (\sigma_2 \circ \Phi^{d_2}(\sigma_1), d_1 \circ d_2)\),
- \(I^\#(\Sigma, i) = I^i(\Sigma)\), and \(I^\#(\sigma, d) = I^i(\Sigma_1) \xrightarrow{\Phi^d} I^i(\Phi^d(\Sigma_1)) \xrightarrow{I^j(\sigma)} I^j(\Sigma_2)\).

Finally, institution comorphism transformations lead to a congruence on Grothendieck signature morphisms: the congruence is generated by

\[
(\theta^\#_{\Sigma}: \Phi^d(\Sigma) \to \Phi^d(\Sigma), d': j \to i) \equiv (id: \Phi^d(\Sigma) \to \Phi^d(\Sigma), d: j \to i)
\]

for \(u : d \Rightarrow d' \in Ind, d, d': j \to i \in Ind, \Sigma \in \text{Sign}_i\).

**Proposition 3.2.** \(\equiv\) is contained in the kernel of \(I^\#\) (considered as a functor).

**Corollary 3.3.** \(I^\#: \text{Sign}^\# \to \text{InsRoom}\) leads to a quotient Grothendieck institution \(I^\# / \equiv: \text{Sign}^\# / \equiv \to \text{InsRoom}\).

\(^2\) More precisely, the corridors provided by these morphisms can be expressed as corridors induced by signature morphisms in the comorphism-based Grothendieck institution.

\(^3\) Strictly speaking, these are not comorphisms, but simple comoidal comorphisms in the sense of [13]. For simplicity, we will ignore this difference here.

\(^4\) \(Ind^*\) is the 2-categorical dual of \(Ind\), where both 1-cells and 2-cells are reversed.
The theory of Grothendieck institutions for indexed institutions has been developed by Diaconescu [8]. Actually, the corresponding theory for indexed coinstitutions turns out to be much simpler. Here, we focus on cocompleteness and amalgamation results, since these are needed for doing structured proofs [18].

**Proposition 3.4.** Let \( I : \text{Ind}^{op} \to \text{colns} \) be an indexed coinstitution such that \( \text{Ind} \) is \( J \)-complete for some small category \( J \), \( \Phi^d \) is cocontinuous for each \( d : i \to j \in \text{Ind} \), and the indexed category of signatures of \( I \) is locally \( J \)-cocomplete (the latter meaning that \( \text{Sign}^i \) is \( J \)-cocomplete for each \( i \in \text{Ind} \)). Then the signature category of the Grothendieck institution has \( J \)-colimits.

Given an institution \( I \) and a diagram \( D : J \to \text{Sign}^i \), a family of models \( (m_j)_{j \in J} \) is called \( D \)-consistent if \( m_k \upharpoonright_{D(k)} = m_j \) for each \( d : j \to k \in J \). A cocone \( (\Sigma, (\mu_j)_{j \in J}) \) over the diagram in \( D : J \to \text{Sign}^i \) is called weakly amalgamable if for each \( D \)-consistent family of models \( (m_j)_{j \in J} \), there is a \( \Sigma \)-model \( m \) with \( m_{\upharpoonright_{D(j)}} = m_j \) (\( j \in |J| \)). If this model is unique, the cocone is called amalgamable.

These notions also extend to diagrams \( D : J \to \text{Ind} \) and cones over these: the rôle of the intra-institution model reductions above is now played by the inter-institution model translations.

An institution \( I \) is called (weakly) semi-exact, if any pullback of signatures is (weakly) amalgamable. An institution comorphism \( (\Phi, \rho) : I_1 \to I_2 \) is called (weakly) exact if, for each signature morphism \( \sigma : \Sigma_1 \to \Sigma_2 \) in \( I_1 \), the naturality diagram

\[
\begin{array}{ccc}
\text{Mod}^i(\Sigma_1) & \xleftarrow{\beta_{\Sigma_1}} & \text{Mod}^i(\Phi(\Sigma_1)) \\
\text{Mod}^i(\Sigma_2) & \xleftarrow{\beta_{\Sigma_2}} & \text{Mod}^i(\Phi(\Sigma_2)) \\
\end{array}
\]

is (weakly) amalgamable, i.e. a (weak) pullback.

An indexed coinstitution \( I : \text{Ind}^{op} \to \text{colns} \) is called (weakly) locally semi-exact, if each institution \( I^i \) is (weakly) semi-exact (\( i \in \text{Ind} \)). It is called (weakly) semi-exact if for each pullback in \( \text{Ind} \)

\[
\begin{array}{cccc}
i & m_1 & j_1 & \text{Mod}^i(\Sigma) \xleftarrow{\beta_{\Sigma}} \text{Mod}^i(\Phi_{m_1}(\Sigma)) \\
m_2 & n_1 & & \text{the square} \xrightarrow{\beta_{\Sigma}} \text{Mod}^i(\Phi_{m_2}(\Sigma)) \xleftarrow{\beta_{\Sigma}} \text{Mod}^i(\Phi_{m_2}(\Sigma)) \\
j_2 & n_2 & k & \text{Mod}^i(\Phi_{m_2}(\Sigma)) \xleftarrow{\beta_{\Sigma}} \text{Mod}^i(\Phi_{m_2}(\Sigma)) \\
\end{array}
\]

is a (weak) pullback for each signature \( \Sigma \) in \( \text{Sign}^i \).

**Proposition 3.5.** The Grothendieck institution \( I^# \) of an indexed coinstitution \( I : \text{Ind}^{op} \to \text{colns} \) consisting of comorphisms with cocontinuous signature translation is (weakly) semi-exact if and only if

- \( I \) is (weakly) locally semi-exact,
- \( I \) is (weakly) semi-exact, and
- all institution comorphisms in \( I \) are (weakly) exact.
Diaconescu proves the above results for so-called embedding-indexed institutions, which means that each signature translation $\Psi^d$ has a left adjoint $\Phi^d$. But these left adjoints lead to a corresponding indexed institution, and in fact, strictly speaking Diaconescu uses this induced indexed institution in his proofs. This shows that indexed co-institutions are simpler and more general than embedding-indexed institutions (and only for these, Diaconescu has results about exactness and amalgamability). In particular, a simpler proof of Diaconescu’s results can be obtained by reducing them to the above results via the following generalization of a result from [1]:

**Proposition 3.6.** Given an embedding-indexed institution $\mathcal{I}: \text{Ind}^{op} \rightarrow \text{Ins}$, define the indexed co-institution $\mathcal{I}^{\text{co}}: \text{Ind}^{op} \rightarrow \text{coIns}$ by

$$\mathcal{I}^{\text{co}}(i) := \mathcal{I}(i) \text{ and } \mathcal{I}^{\text{co}}(d;i \rightarrow j) := (\Phi^d, (\mu^d \cdot \Phi^d) \circ (\mathcal{I}^j \cdot \eta^d)),$$

where $\eta^d$ is the unit of the adjunction between $\Phi^d$ and $\Psi^d$.

Then $\mathcal{I}^\#$ is isomorphic to $(\mathcal{I}^{\text{co}})^\#$.

Diaconescu already notes that the assumptions of (his more special version of) Prop. 3.5 are too strong to be met in practice. E.g. the CAS1 institution is not weakly locally semi-exact, and its encoding into HOL‘ is neither exact, nor does it have a cocontinuous signature translation. Below, we will try to weaken the assumptions by working with weakly amalgamable cones rather than with amalgamable colimits: call an institution $\mathcal{I}$ quasi-exact if for each diagram $D: J \rightarrow \text{Sign}^I_j$, there is some weakly amalgamable cocone over $D$. An indexed co-institution $\mathcal{I}: \text{Ind}^{op} \rightarrow \text{coIns}$ is called locally quasi-exact, if each institution $\mathcal{I}^i$ is quasi-exact ($i \in |\text{Ind}|$). It is called quasi-exact, if for each diagram $D: J \rightarrow \text{Ind}$, there is some cone $(l_i(j), j \in |J|)$ over $D$ that is weakly amalgamable. **Quasi-semi-exactness** is the restriction of these notions to diagrams of shape

```
• ←• →•
```

4 Heterogeneous development graphs

We now come to proof theory:

**Proposition 4.1.** The Grothendieck logic $L^\#$ of an indexed co-logic $\mathcal{L}: \text{Ind}^{op} \rightarrow \text{coLog}$ is complete if and only if $\mathcal{L}$ is locally complete (i.e. each individual logic is complete).

In many cases, not every institution will come with a complete entailment system; hence, it is difficult to apply the above proposition in practice. In [17], we therefore have represented $L^\#$ in some expressive “universal” institution with good proof support. However, often it is crucial to use specific tools designed for specific logics to obtain good results. Hence, we need heterogeneous proving as well. A first attempt in this direction are heterogeneous bridges [7] which, however, have no clear semantical basis. We therefore aim at clear semantical basis for heterogeneous specification and proofs, where in the extreme for each proof
goal the best-suited logic could be chosen individually. For this end, we need some heterogeneous structuring concept, namely heterogeneous development graphs.

In [17], we have defined heterogeneous development graphs to be homogeneous development graphs over a Grothendieck institution. We here directly define development graphs and their proof calculus over a Grothendieck institution, since this allows us to design a specialized rule called Model expansion, which together with a rule Borrowing allows us to deal with the re-use of entailment relations across comorphisms [6].

Fix an arbitrary 2-indexed constitution \( I: \text{Ind}^* \rightarrow \text{colIns} \). Let \( (\text{Sign}^\#, \equiv, \text{Sen}^\#, \equiv, \text{Mod}^\#, \equiv, \equiv^\#) \) denote the components of the quotient Grothendieck institution \( I^\# \equiv \). We need the congruence \( \equiv \) induced by the 2-categorical structure (and the corresponding quotient), because with it we can prove more things in the calculus introduced below.

We further assume that some of the institutions come as logics. Moreover, we assume that some of the arrows \( d \in \text{Ind} \) are marked with the fact that \( I^d \) has the model expansion property (i.e. the model translation components of the corridors are surjective on objects). Finally, we assume that there is a computable partial function yielding a weakly associative (co)cone for some of the diagrams in \( \text{Sign}^\# \equiv \). These assumptions are reasonable in practice.

**Definition 4.2.** A heterogeneous development graph (over \( I \)) is an acyclic directed graph \( S = (N, L) \).

\( N \) is a set of nodes. Each node \( N \in N \) is a tuple \((\Sigma^N, \Gamma^N)\) such that \( \Sigma^N \in \text{Sign}^\# \equiv \) is a signature and \( \Gamma^N \subseteq \text{Sen}^\# \equiv (\Sigma^N) \) is the set of local axioms of \( N \).

\( L \) is a set of directed links, so-called definition links, between elements of \( N \). Each definition link from a node \( M \) to a node \( N \) is either

- global (denoted \( M \xrightarrow{\sigma} N \)), annotated with a signature morphism \( \sigma : \Sigma^M \rightarrow \Sigma^N \in \text{Sign}^\# \equiv \), or
- hiding (denoted \( M \xrightarrow{\sigma} \overline{N} \)), annotated with a signature morphism \( \sigma : \Sigma^N \rightarrow \Sigma^M \in \text{Sign}^\# \equiv \) going against the direction of the link. Typically, \( \sigma \) will be an inclusion, and the symbols of \( \Sigma^M \) not in \( \Sigma^N \) will be hidden.

**Definition 4.3.** Given a node \( N \in N \), its associated class \( \text{Mod}_S(N) \) of models (or \( N \)-models for short) consists of those \( \Sigma^N \)-models \( n \) for which

- \( n \) satisfies the local axioms \( \Gamma^N \),
- for each \( K \xrightarrow{\sigma} N \in S, n|_{\sigma} \) is a \( K \)-model, and
- for each \( K \xrightarrow{\sigma} N \in S, n \) has a \( \sigma \)-expansion \( k \) (i.e. \( k|_{\sigma} = n \)) which is a \( K \)-model.

The notion of global reachability is defined inductively: A node \( M \) is globally reachable from a node \( N \) via a signature morphism \( \sigma \), \( N \xrightarrow{\sigma} M \) for short,
iff either $N = M$ and $\sigma = id$, or $N \xrightarrow{\sigma'} K \in S$, and $K \xrightarrow{\sigma''} M$, with $\sigma = \sigma'' \circ \sigma'$.

Complementary to definition and hiding links, which define the theories of related nodes, we introduce the notion of a theorem link with the help of which we are able to postulate relations between different theories. Global and theorem links (denoted by $N \xrightarrow{\sigma} M$ and $N \xrightarrow{\sigma} M$ respectively, where $\sigma: \Sigma^N \rightarrow \Sigma^M$) are the central data structure to represent proof obligations arising in formal developments. We also need theorem links $N \xrightarrow{\sigma} M$ (where for some $\Sigma$, $\theta: \Sigma \rightarrow \Sigma^N$ and $\sigma: \Sigma \rightarrow \Sigma^M$) involving hiding.

**Definition 4.4.** Let $S$ be a development graph. $S$ implies a global theorem link $N \xrightarrow{\sigma} M$ (denoted $S \models N \xrightarrow{\sigma} M$), iff for all $m \in \text{Mod}_S(M)$, $m|_{\sigma} \in \text{Mod}_S(N)$. $S$ implies a local theorem link $N \xrightarrow{\sigma} M$, if for all $m \in \text{Mod}_S(M)$, $m|_{\sigma} \models \Gamma^N$. Finally, $S$ implies a hiding theorem link $N \xrightarrow{\sigma_h} M$, iff for all $m \in \text{Mod}_S(M)$, $m|_{\sigma}$ has a $\theta$-extension to an $N$-model.

A global definition link $M \xrightarrow{\sigma} N$ in a development graph is a conservative extension if every $M$-model can be expanded along $\sigma$ to an $N$-model. We will allow annotating a global definition link as $M \xrightarrow{\sigma} N$, which expresses that it is a conservative extension. These annotations can be seen as another kind of proof obligations.

There are quite a number of institution independent languages for structured specifications [21,12,10,15,11,19], one of which also has been extended to the heterogeneous case [23]. Most of their constructs can be translated into the formalism of development graphs, which hence can be seen as a core formalism for structured and heterogeneous theorem proving. For the language Castl, such a translation has been laid out explicitly in [2]. An exception are freeness constraints, which are currently not present in development graphs, because a logic-independent proof theory for them is missing (yet feasible at all). We now extend our proof calculus for development graphs from [18] to the heterogeneous case. The rules are typically applied backwards, thereby possibly adding some new nodes and edges to the development graph.

The central rule of the proof system is the rule Theorem-Hide-Shift (cf. Fig. 2). It is used to get rid off hiding definition links going into the target of a global theorem link. Since it is quite powerful, we need some preliminary notions. Given a node $N$ in a development graph $S = \langle N, L \rangle$, the idea is that we unfold the subgraph below $N$ into a tree and form a diagram with this tree. More formally, define the diagram $D: J \rightarrow \text{Sign associated with } N$ together with a map $G: [J] \rightarrow N$ inductively as follows:

- $\langle N \rangle$ is an object in $J$, with $D(\langle N \rangle) = \Sigma^N$. Let $G(\langle N \rangle)$ be just $N$.
- if $i = \langle M \xrightarrow{l_1} \cdots \xrightarrow{l_n} N \rangle$ is an object in $J$ with $l_1, \ldots, l_n$ definition links in $L$, and $l = K \xrightarrow{\sigma} M$ is a global definition link in $L$, then
\[ j = \langle K \xrightarrow{L} M \xrightarrow{l_1} \cdots \xrightarrow{l_n} N \rangle \]
is an object in \( J \) with \( D(j) = \Sigma^K \), and \( l \) is a morphism from \( j \) to \( i \) in \( J \) with \( D(l) = \sigma \). We set \( G(j) = K \).

- if \( i = \langle M \xrightarrow{l_1} \cdots \xrightarrow{l_n} N \rangle \) is an object in \( J \) with \( l_1, \ldots, l_n \) definition links in \( \mathcal{L} \), and \( l = K \xrightarrow{\alpha} M \) is a hiding definition link in \( \mathcal{L} \), then

\[ j = \langle K \xrightarrow{L} M \xrightarrow{l_1} \cdots \xrightarrow{l_n} N \rangle \]
is an object in \( J \) with \( D(j) = \Sigma^K \), and \( l \) is a morphism from \( i \) to \( j \) in \( J \) with \( D(l) = \sigma \). We set \( G(j) = K \).

Now in order to apply Theorem-Hide-Shift, \((\Sigma, (\mu_i : D(i) \to \Sigma)_{i \in |J|})\) has to be a weakly amalgamable cocone for \( D \), and \( C \) has to be a new isolated node with signature \( \Sigma \) and with ingoing global definition links \( G(i) \xrightarrow{\mu_i} C \) for \( i \in |J| \).

Here, an isolated node is one with no local axioms and no ingoing definition links other than those shown in the rule.

\[
\text{Theorem-Hide-Shift}
\]

\[
K \xrightarrow{\sigma \alpha} M \quad \text{for each} \quad K \xrightarrow{\sigma'} N \]
\[
L \xrightarrow{\sigma \alpha} M \quad \text{for each} \quad L \xrightarrow{\theta} K \quad \text{and} \quad K \xrightarrow{\sigma'} N
\]

\[ N \xrightarrow{\theta} M \]

\[ \text{Glob-Decomposition} \]

Fig. 2. Structural rules

In order to get rid of hiding links going into the source of a global theorem link, one first applies Glob-Decomposition, ending up with some local and hiding
Theorem links. The rule Hide-Theorem-Shift allows to prove the latter, using conservativeness of definition links. Borrowing is mainly used for shifting a proof goal into a different logic; it also exploits conservativity of definition links. We therefore also need rules dealing with conservativity:

<table>
<thead>
<tr>
<th>Cons-Shift</th>
<th>Cons-Composition</th>
<th>Model-Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M \xrightarrow{\sigma} N$ if $\Sigma^M \xrightarrow{\sigma} \Sigma^N$</td>
<td>$M \xrightarrow{\sigma} N$ if $d$ is marked as model-expansive and $N$ is isolated.</td>
<td>$M \xrightarrow{c} N$</td>
</tr>
<tr>
<td>$M' \xrightarrow{\sigma} N'$ is weakly amalgamable and $N'$ is isolated.</td>
<td>$M \xrightarrow{\sigma} N$ if $d \in (a, d)$</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 3.** Conservativity rules

Finally, we have a set of decomposition rules not interacting with hiding, and a rule Basic Inference allowing to reduce local theorem links to inference in the calculus of some of the logics in the indexed coinstitution:

**Subsumption:**

$\frac{N \xrightarrow{\sigma} M}{N \xrightarrow{\sigma} M}$

**Loc-Decomposition I:**

$\frac{K \xrightarrow{\sigma} L \quad \text{if } L \xrightarrow{\sigma'} M \text{ and } \sigma''(\Gamma^K) = \sigma'((\Gamma^K))}{K \xrightarrow{\sigma''} M}$

**Loc-Decomposition II:**

$\frac{N \xrightarrow{\sigma'} M \quad \text{if } \sigma(\Gamma^N) = \sigma'(\Gamma^N)}{N \xrightarrow{\sigma} M}$

**Basic Inference:**

$\frac{\text{Th}_S(M) \vdash_{\Sigma^M} \sigma(\varphi) \text{ for each } \varphi \in \Gamma^N}{N \xrightarrow{\sigma} M}$

Here, $\text{Th}_S(M)$ is inductively defined to be

$\Gamma^M \cup \bigcup_{K \xrightarrow{\sigma} M \in S} \sigma(\text{Th}_S(K))$
This is well-defined because development graphs have to be acyclic.

**Theorem 4.5.** For a 2-indexed constituteion $\mathcal{I}: \text{Ind}^* \rightarrow \text{cols}$ (some of which come as logics), the proof calculus for heterogeneous development graphs is sound for $\mathcal{I}^#/\equiv$. If, moreover,

- $\mathcal{I}$ is quasi-exact,
- all institutions comorphisms in $\mathcal{I}$ are weakly exact and model-expansive,
- there is a set $\mathcal{L}$ of institutions in $\mathcal{I}$ that come as complete logics,
- the rule system is extended with a (sound and complete) oracle for conservative extension for each logic in $\mathcal{L}$,
- all institutions in $\mathcal{L}$ are quasi-semi-exact,
- from each institution in $\mathcal{I}$, there is some comorphism in $\mathcal{I}$ going into some logic in $\mathcal{L}$, and
- hiding links are only used with signature morphisms whose comorphism component is model-bijective (i.e. the model translation is bijective on objects),

then the proof calculus for heterogeneous development graphs is sound and complete for $\mathcal{I}^#/\equiv$.

Note that due to the Gödel incompleteness theorem, one cannot expect to drop the oracle for conservative extensions, see [18]. The crucial achievement here is to restrict the oracle to intra-logic conservativity.

Further note that in contrast to Prop. 3.5, we need neither cocontinuity nor exactness of the comorphism signature translations here. Moreover, we need quasi-exactness only for some of the logics; this allows us to include logics which are not quasi-exact, such as CASL.

## 5 Conclusion and related work

We have dualized the Grothendieck institution approach to heterogeneous specification to the case of institution comorphisms, which leads to simpler and more general results concerning exactness and weak amalgamation.

We have extended the proof calculus for development graphs with hiding to this setting, and we have studied conditions for its soundness and completeness, which are related to various forms of exactness and weak amalgamation conditions. These conditions are so mild that they hold in typical practical examples; in particular, they are considerably weaker than both the exactness conditions for Grothendieck institutions in [8] and the Craig interpolation property needed for completeness of calculi for structured specification [5]. MAYA [3] implements development graphs and the calculus, but without heterogeneity yet.

**Acknowledgments** Thanks to Andrzej Tarlecki, Serge Auterier and Dieter Hutter for useful cooperation and discussions, and to Lutz Schröder and the anonymous referees for valuable hints. This work has been supported by the project MULTIPLE of the Deutsche Forschungsgemeinschaft under Grant KR 1191/5-1.
References

A Appendix: Proofs of the results

Proof of Proposition 3.2: By the definition of comorphism transformation, \((I_2 \cdot \theta) \circ \rho = \rho'\). But this just means that equivalent signature morphisms induce the same corridors.

Corollary 3.3 follows immediately. \(\Box\)

Proof of Proposition 3.4: Apply Corollary 3 of [13] with \(C_i = \text{Sign}^i\) and \(C_m = \Phi^m\). Note that \(\text{Sign}^\#\) is then \(\text{Flat}(C^\#)^\text{op}\). \(\Box\)

Proof of Proposition 3.5: This follows much in the same way as the corresponding result in [8]. \(\Box\)

Proof of Proposition 3.6: Let \(\eta^d\) be the unit and \(\varepsilon^d\) the counit of the adjunction between \(\Psi^d\) and \(\Phi^d\). Define an institution comorphism \((\Phi, \rho): I^\# \rightarrow (I^\#)^\#\) as follows: \(\Phi\) sends \((\Sigma, i)\) to \((\Sigma, i)\) and \((\sigma, d): (\Sigma_1, i) \rightarrow (\Sigma_2, j)\) in \(I^\#\) to \((\varepsilon_{\Sigma_2} \circ \Phi^d(\sigma), d): (\Sigma_1, i) \rightarrow (\Sigma_2, j)\) in \((I^\#)^\#\). Since \(\varepsilon_{\Sigma_2} \circ \Phi^d(\sigma)\) is just the arrow adjoint to \(\sigma\), \(\Phi\) is an isomorphism.

Now \(\eta^d(\sigma, d)\) is \(I^i(\Sigma) \xrightarrow{I^i(\sigma)} I^i(\Psi^d(\Sigma_2)) \xrightarrow{\mu^d_{\Sigma_2}} I^i(\Sigma_2)\). Since

\[ I^\#(d: j \rightarrow i) = (\Phi^d, (\mu^d \cdot \Phi^d) \circ (I^i \cdot \eta^d)) \]

we get that \((I^\#)^\#(\Phi(\sigma, d)) = (I^\#)^\#(\varepsilon_{\Sigma_2} \circ \Phi^d(\sigma), d) = I^i(\Sigma_1) \xrightarrow{\eta^d_{\Sigma_1}} I^i(\Psi^d(\Sigma_2)) \xrightarrow{\mu^d_{\Sigma_2}} I^i(\Sigma_2)\).

By the following diagram, both are the same, showing that \(\rho\) can be taken to be the identity:

The squares commute by naturality of \(\eta^d\) and \(\mu^d\), while the triangle commutes by a general adjointness law. \(\Box\)
Proof of Proposition 4.1: The logical rooms of the Grothendieck logic \( \mathcal{L}^\# \) are just logical rooms of some individual logic in \( \mathcal{L} \). Hence, global completeness is equivalent to local completeness. \( \square \)

Proof of Theorem 4.5:

\textbf{Soundness:}

The soundness of most of the rules follows from the soundness proof in [18], since we always have assumed the weak amalgamabilities that are needed in [18]. There are three new rules, which we now prove to be sound:

\textit{Borrowing:} Assume that \( S \models M' \models N' \). We need to show \( S \models M \models N \).

Let therefore \( n \) be an \( N \)-model. Since the global definition link \( n \models \theta' \) is conservative, \( n \) can be expanded to an \( N' \)-model \( n' \). By the assumption, \( n'|_{\sigma} \) is then an \( M' \)-model, hence, \( n'|_{\sigma \circ \theta} = n'|_{\theta} = n \) is an \( M \)-model.

\textit{Cons-Composition:} By assumption, any \( M \)-model can be \( \sigma \)-expanded to an \( N \)-model, which in turn can be \( \theta \)-expanded to an \( N' \)-model. Hence, each \( M \)-model can be \( \theta \circ \sigma \)-expanded to an \( N' \)-model.

\textit{Model-Expansion:} By assumption, \( I^d \) is model-expansive. But \( \rho^d \) is just \( I^d(id_d, d) \).

Since \( N \) is isolated, any \( M \)-model can be \( (id_d, d) \)-expanded to an \( N \)-model.

\textbf{Completeness:}

We first need some preparatory lemmas:

\textbf{Lemma A.1.} Let \( I \) satisfy the assumptions of the completeness theorem (Theorem 4.5). Then \( I^\# \) is quasi-semi-exact.

Proof. Let a diagram \( (\Sigma_1, j_1) \xrightarrow{(\sigma_1, m_1)} (\Sigma, i) \xrightarrow{(\sigma_2, m_2)} (\Sigma, j_2) \) in \( \text{Sign}^\# \) be given. Let

\[
\begin{array}{c}
\Sigma_1 \xrightarrow{j_1} \Sigma \\
\downarrow \sigma_1 \downarrow \downarrow \downarrow \\
\Sigma_2 \xrightarrow{j_2} \Sigma
\end{array}
\]

be a weakly amalgamable cone over \( Ind \). Let \( n: k \rightarrow k' \) be such that \( I^n \) is a logic in \( \mathcal{L} \) (i.e. also being quasi-semi-exact). Let \( n_1 := n_1 \circ n \) and \( n_2 := n_2 \circ n \).

Since \( I^n \) is model-expansive,

\[
\begin{array}{c}
\Sigma_1 \xrightarrow{j_1} \Sigma \\
\downarrow \sigma_1 \downarrow \downarrow \downarrow \\
\Sigma_2 \xrightarrow{j_2} \Sigma
\end{array}
\]
is weakly amalgamable as well (leading to weak amalgamability of the upper left square below). Let the lower right square of

\[
\begin{array}{c}
(\Sigma, i) \xrightarrow{\text{id}, m_2} (\phi_{\text{id}}(\Sigma), j_2) \xrightarrow{\text{id}, \text{id}} (\phi_{\text{id}}(\Sigma), j_2) \\
\downarrow \phi_{\text{id}} \quad \downarrow \phi_{\text{id}} \quad \downarrow \phi_{\text{id}} \quad \downarrow \phi_{\text{id}} \\
(\Sigma_1, j_1) \xrightarrow{\text{id}, m_1} (\phi_{\text{id}}(\Sigma_1), j_1) \xrightarrow{\text{id}, \text{id}} (\phi_{\text{id}}(\Sigma_1), j_1)
\end{array}
\]

be a weakly amalgamable cone in $\text{Sign}^R$ (existing by quasi-semi-exactness of $k$). By weak exactness of $\mathcal{I}^{n_1}$ and $\mathcal{I}^{n_2}$, also the remaining two squares are weakly amalgamable. Since weakly amalgamable squares can be pasted together, we get a weakly amalgamable cone for the original diagram. □

**Lemma A.2.** Let $\mathcal{I}$ satisfy the assumptions of the completeness theorem (Theorem 4.5). Then $\mathcal{I}^R$ admits weak amalgamability of acyclic connected diagrams (here, acyclicity of course disregards the identities).

Proof. In the sequel, we will use terms like “connected”, “maximal”, “lower bound” for small categories, when we really mean the pre-order obtained from the category by collapsing the hom-sets into singletons.

Let $D: J \rightarrow \text{Sign}^R$ be a connected diagram and let $\text{Max}$ be the set of maximal nodes in $J$. We successively construct new diagrams out of $J$. Take two nodes in $\text{Max}$ that have a common lower bound (if such two nodes do not exist, the diagram is not connected). Take a weak amalgamating cone for the sub-diagram consisting of the two maximal nodes and the lower bound (together with the arrows from the lower bound into the maximal nodes). Extend the diagram with the cone. The thus obtained diagram now has a set $\text{Max}$ of maximal nodes that is decreased by one. By iterating this construction, we get a diagram with one maximal node. The maximal node then is just the tip of a weakly amalgamating cone for the original diagram. □

The next lemma is proved in [18]:

**Lemma A.3.** If $C$ is constructed as in the rule *Theorem-Hide-Shift*, then any $\Sigma^C$-model satisfying $\text{Th}_{\mathcal{C}}(C)$ is already a $C$-model.

We now come to the proof of the completeness theorem.

Assume $\mathcal{S} \models M \not\not\not\not\not\not_{\mathcal{C}} \not\not\not\not\not\not N$. We show that there is some faithful extension $\mathcal{S}_1$ of $\mathcal{S}$ (i.e. new nodes and new definition links are added, but the latter go only into new nodes) such that $\mathcal{S}_1 \models M \not\not\not\not\not\not_{\mathcal{C}} \not\not\not\not\not\not N$.

Let $D: J \rightarrow \text{Sign}$ and $C$ be as in the rule *Theorem-Hide-Shift* (note that by Lemma A.2, a weakly amalgamable cone exists, since the diagram constructed in the rule is acyclic and connected). Let $(\Sigma^C, \Sigma^C)$ be the signature of $C$, and
let $c$ be a $(\Sigma^C, i^C)$-model satisfying $\text{Th}_S(C)$. By Lemma A.3, $c$ is a $C$-model. Hence, $c|_{\mu(N)}$ is an $N$-model, and by the assumption $S \models M \rightarrow C \rightarrow N$, $c|_{\mu(N) \circ \sigma}$ is an $M$-model. We now have for any $K \rightarrow M$:

1. $c|_{\mu(N) \circ \sigma \circ \theta} \models \Phi^K$. By the satisfaction condition for the Grothendieck institution $I^\#$, we get $c \models \mu(N)(\sigma(\theta(\Phi^K)))$. Hence, we have shown $\text{Th}_S(C) \models \mu(N)(\sigma(\theta(\Phi^K)))$. Let $d: i \rightarrow i$ be such that $I^d$ is a comorphism from $I^i$ to $I^i$, where the latter also is a complete logic (this exists by assumption of the theorem). Obtain a new development graph $S'$ from $S$ by letting $C'$ be a new node with signature $(\Phi^d(\Sigma^C), I)$ and with one ingoing definition link $C \xrightarrow{(id,d)} C'$.

\[ K \xrightarrow{\mu(N) \circ \sigma \circ \theta} C \]

\[ \xrightarrow{id} \]

\[ \xrightarrow{(id,d) \circ \mu(N) \circ \sigma \circ \theta} C' \]

By the satisfaction condition for comorphism corridors, we get $\text{Th}_S(C') \models \alpha^K(\mu(N)(\sigma(\theta(\Phi^K))))$. By completeness of the logic $I^d$, we obtain $\text{Th}_S(C') \models \alpha^K(\mu(N)(\sigma(\theta(\Phi^K))))$. By Basic Inference, $S \models K \rightarrow C'$. Since by assumption, all comorphisms are model-expansive, by Model-Expansion, $(id,d)$ is conservative. By Borrowing, $S \models K \rightarrow C$.

2. For $L \xrightarrow{(\sigma_2, m_2)} K$, by Lemma A.1, we obtain a weakly amalgamating cocone

\[ \xymatrix{ \Sigma^K \ar[r]_{\mu(N) \circ \sigma \circ \theta} & \Sigma^C \ar[d]^{(\sigma_1, m_1)} \ar[d]_{(\theta_1, m_1)} \ar[r]^{(\sigma_2, m_2)} & \Sigma^L \ar[r]^{(\theta_2, m_2)} & (\Sigma', k) } \]

By inspecting the proof of Lemma A.1, we can assume that the above diagram is split up into a diagram of four squares, and that $I^k$ that is in the set $L$ of complete logics. We will interchangeably also use the notation of the diagram in the proof of Lemma A.1, i.e. we put $(\sigma_1, m_1) := \mu(N) \circ \sigma \circ \theta$ and $(\Sigma_1, j_1) := (\Sigma^C, i^C)$.

We now construct a new development graph $S'$ from $S$ by introducing a new node $L'$ with signature $(\Sigma', k)$, a new node $C'$ with signature $(\Phi^{m_1}(\Sigma_1), k)$. $C'$ has one ingoing definition link $C \xrightarrow{(id, m_1)} C'$, while $L'$ has two ingoing definition links $L \xrightarrow{(\theta_2, m_2)} L'$ and $C \xrightarrow{(\theta_1, id)} L'$. The latter link is conservative, which can be seen as follows: for any $C'$-model $c'$, $c'|_{(id, m_1) \circ (\sigma_1, m_1)}$...
has a \((\sigma_2, m_2)\)-expansion to an \(L\)-model \(c_2\). Since by assumption, hiding is done only against signature morphisms with model-bijection, \(L^{m_2}\) is model-bijective, and we get \(c' \models (\phi, \Sigma_1, n_1) = c_2 \models (\sigma_2, \text{id})\). By weak amalgamation, we get a \((\Sigma', k)\)-model \(c_3\) with \(c_3 \models (\sigma_1, \text{id}) = c_1\) and \(c_3 \models (\theta_2, n_2) = c_2\). These properties show \(c_3\) to be an \(L'\)-model. By the oracle for conservativity in logic \(T^k\) (note that by assumption all logics in \(L\) come with such an oracle), we get \(C \models (\theta_1, n_1) \to \to L'\). By assumption, all monomorphisms are model-expansive, and by Model-Expansion, \(C \models (\text{id}, n_1) \to \to C'\). With Cons-Composition, we get \(C \models (\theta_1, n_1) \to \to L'\). Now \(S' \models L \to \to L'\) by Subsumption.

By Hide-Theorem-Shift, we get \(S' \models L \to \to C\).

Let \(S_1\) be the union of all the \(S'\) constructed in steps 1 and 2 above (assuming that all the added nodes are distinct). By Glob-Decomposition, we get \(S_1 \models M \to \to C\). By Theorem-Hide-Shift, we get \(S_1 \models M \to \to N\). \(\square\)